Methods xxx (xxxx) xxx

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Radiomics for precision medicine: Current challenges, future prospects, and the proposal of a new framework

A. Ibrahim [a,b,c,d,*], S. Primakov [a,d,1], M. Beuque [a,1], H.C. Woodruff [a,b], I. Halilaj [a], G. Wu [a], T. Refaee [a,e], R. Granzier [b,f], Y. Widaatalla [a], R. Hustinx [c], F.M. Mottaghy [b,d], P. Lambin [a,b]

[a] The D-Lab, Department of Precision Medicine, GROW – School for Oncology and Developmental Biology, Maastricht University, Maastricht, The Netherlands
[b] Department of Radiology and Nuclear Medicine, GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht, The Netherlands
[c] Division of Nuclear Medicine and Oncological Imaging, Department of Medical Physics, Hospital Center Universitaire De Liege, Liege, Belgium
[d] Department of Nuclear Medicine and Comprehensive Diagnostic Center Aachen (CDCA), University Hospital RWTH Aachen University, Aachen, Germany
[e] Department of Diagnostic Radiology, Faculty of Applied Medical Sciences, Jazan University, Jazan, Saudi Arabia
[f] Department of Surgery, GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands

ARTICLE INFO

Keywords:
Radiomics
Clinical decision support systems
Medical image analysis

ABSTRACT

The advancement of artificial intelligence concurrent with the development of medical imaging techniques provided a unique opportunity to turn medical imaging from mostly qualitative, to further quantitative and mineable data that can be explored for the development of clinical decision support systems (cDSS). Radiomics, a method for the high throughput extraction of hand-crafted features from medical images, and deep learning -the data driven modeling techniques based on the principles of simplified brain neuron interactions, are the most researched quantitative imaging techniques. Many studies reported on the potential of such techniques in the context of cDSS. Such techniques could be highly appealing due to the reuse of existing data, automation of clinical workflows, minimal invasiveness, three-dimensional volumetric characterization, and the promise of high accuracy and reproducibility of results and cost-effectiveness. Nevertheless, there are several challenges that quantitative imaging techniques face, and need to be addressed before the translation to clinical use. These challenges include, but are not limited to, the explainability of the models, the reproducibility of the quantitative imaging features, and their sensitivity to variations in image acquisition and reconstruction parameters. In this narrative review, we report on the status of quantitative medical image analysis using radiomics and deep learning, the challenges the field is facing, propose a framework for robust radiomics analysis, and discuss future prospects.

1. Introduction

Advances in artificial intelligence applications, combined with those in medical imaging, have led to the gradual conversion of digital medical images into high-dimensional data appropriate for data mining and data science techniques [1]. Meanwhile, computing power and quantitative image analysis (QIA) techniques have made enormous progress, and the application of quantitative imaging techniques on medical imaging gained exponential momentum [2]. Currently, radiomics and deep learning are the most researched techniques on medical imaging.

Broadly, radiomics refers to the use of computational or statistical approaches to extract large numbers of quantitative features from a number of medical imaging modalities, such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET), to develop predictive models ultimately aiming to enable personalized clinical management [3–5]. Radiomic features are quantitative descriptions of the intensity, shape, volume, and texture of the region of interest (ROI), with the recent addition of more abstract features such as radial gradient and radial deviation [6]. Radiomics features are broadly divided into histogram-based and texture features. Different statistical methods are used to calculate the radiomics features. The methods include first-order statistics, which depends on the values of single voxels (histogram-based features for e.g. maximum and minimum intensity); second-order statistics, which depends on the relation

* Corresponding author at: Precision Medicine–UM, P.O.B 612, 6200MD Maastricht, The Netherlands.
E-mail address: a.ibrahim@maastrichtuniversity.nl (A. Ibrahim).
[1] Both authors contributed equally.

https://doi.org/10.1016/j.ymeth.2020.05.022
Received 2 April 2020; Received in revised form 20 May 2020; Accepted 26 May 2020
Available online 3 June 2020

between two voxels (for e.g. grey-level co-occurrence matrix (GLCM) features), and higher-order statistics (relations among three or more voxels, for e.g. neighborhood grey-tone difference matrices (NGTDM) features) [7,8]. The main hypothesis behind radiomics analysis is that radiomic features decode or correlate with the molecular characteristics, phenotype, and genotype of the region of interest (ROI) under study. This information can be used in combination with other patient information to improve patient management. Moreover, as the tumors are of heterogeneous nature [9,10], clinical approaches, such as tissue biopsies, might fail to characterize the entirety of the tumor [11]. In contrast, Radiomics takes the whole tumor region (or even the surrounding or healthy tissue) into account, which enables a better characterization [3]. Furthermore, frequent clinical imaging can transform radiomics into a non-invasive, easily repeatable, and cost-effective longitudinal approach for cDSS [12].

Deep learning (DL) is a field of data driven modelling techniques that utilizes the principles of simplified neuron interactions [13]. Using artificial neurons started to draw attention decades ago [14], but it only became a major research focus recently [15–17]. The artificial neuron model is used as a foundation unit to create complex chains of interactions – DL layers. These layers are used to generate even more complex structures - DL architectures (see Fig. 1). The neural network (NN) training procedure is typically a cost-function minimization process. The cost function measures the error of predictions based on the ground truth labels [18]. Due to the high complexity of the network architectures, computational limitations are reached when trying to solve the optimization task analytically. Henceforth, iterative algorithms are used to overcome this issue. Commonly, these algorithms are variations of the gradient descent (GD). GD iteratively moves in the direction of steepest descent of the cost function, in order to find a local minimum. During the model training process, every image from the training dataset contributes to the cost minimization process. Thereby, a DL network learns how to solve a problem directly from existing data, and apply it to data it has never seen. These complex models contain the parameters (weights) for millions of neurons, which can be trained for the recognition of problem-related patterns in the data being analyzed. DL has been shown to be efficient in other fields, such as face recognition [19] and autonomous cars [20].

Since the introduction of the field, many studies have reported on the potential of such techniques for predicting patient outcomes [5,21,22]. The successful translation of QIA techniques into cDSS will have a significant impact on the clinical workflow and current patient management protocols. Clinicians will be able to non-invasively obtain a more detailed and accurate tumor characterization, in a shorter amount of time. Patients will have to go through less invasive procedures, while having treatment optimized based on their individual characteristics. Furthermore, patient-specific informed decisions can be made with more confidence. However, QIA is still developing in the field of medical imaging and several challenges, including the stability and reproducibility of imaging biomarkers, as well as the interpretability of the developed algorithms, need to be addressed before QIA can be translated to clinical applications.

In this narrative review, we focus on the current status of the potential of radiomics and deep learning to be incorporated in clinical decision support systems (cDSS), their challenges, as well as future prospects for these methods. We further propose a workflow to guide robust radiomics analysis.

2. Quantitative image analysis for precision medicine

The need for personalizing the management of patients has been widely reported [23,24]. QIA represents a suitable candidate to be incorporated into the body of personalized medicine due to the non-invasive three-dimensional characterization of the ROIs, the availability of vast amounts of medical images, the longitudinal capabilities, and the cost-effectiveness of the method.

The currently implemented imaging biomarker development workflow is generalizable across different imaging modalities. The workflow can be described as consecutive steps divided into the main categories of data collection, image segmentation, features extraction, development of the signature, and evaluation of the performance (Fig. 2), with the segmentation step being optional in the case of deep learning. The workflow has been previously extensively described [22,25].

Many studies have investigated and reported on the added clinical value of radiomics features for predicting various clinical outcomes, such as overall survival, tumor histology, response to therapy, and genetic profiling, among other endpoints. Furthermore, these studies were performed on various imaging modalities, including CT, MR, and PET.

While the hand-crafted radiomics pipeline necessitates the use of machine learning or statistical algorithms after feature extraction for modeling, DL techniques perform feature extraction and modelling internally without the need for further user interaction. DL has its own advantages and drawbacks compared to traditional radiomics. One of the key benefits of using DL is avoiding the contouring problem, the bottleneck of a traditional radiomics pipeline. However, due to the complexity of DL models, it is easier to overfit the model to the training data. As a result, a larger data set is needed for DL compared to hand-crafted radiomics. Furthermore, DL is considered a 'black box', i.e the models and features generated are not (or barely) interpretable. This is currently one of the major challenges of the application of artificial intelligence (AI) in medical image analysis. Efforts are being made towards providing explainable AI algorithms, by investigating the correlation of the chosen features with biologic or semantic characteristics. Such correlations would provide an understanding about how the algorithm makes the decision, and ease its incorporation into cDSS.

QIA techniques have a great potential for involvement in developing classification, prognostic and predictive clinical tools. In comparison, classification tasks (for e.g classifying tissue histology) seem to yield a better performance than predictive tasks (for e.g survival prediction). This is in part due to the unaccounted for variables when trying to predict future events. In 2.1 and 2.2, we report on some examples that highlighted the potential of radiomics and deep learning to predict various clinical endpoints, acknowledged or addressed the challenges of QIA techniques used, and/or applied the techniques on a relatively large sample size compared to other studies addressing the same clinical endpoint.

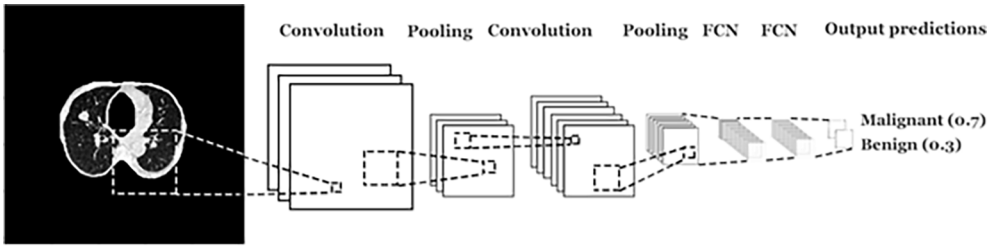

Fig. 1. Graphical depiction of DL architectures. * FCN: fully connected network.

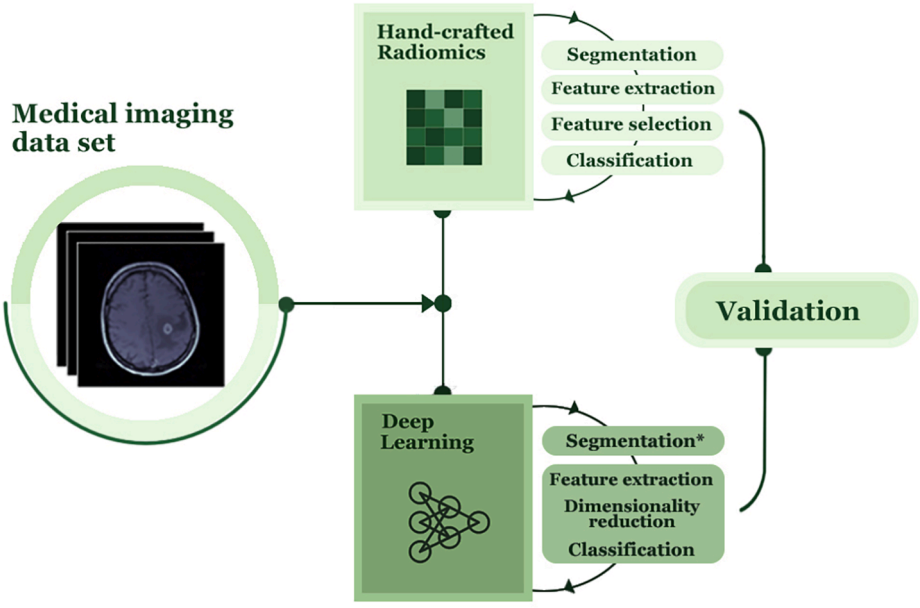

Fig. 2. Development of imaging biomarkers using quantitative image analysis. * Segmentation is not a necessity in the automated radiomics pipeline.

2.1. Hand-crafted radiomics

2.1.1. Overall survival

Wang et al. [26] investigated the potential of radiomics signatures to predict overall survival in patients with locally advanced rectal cancer. The authors tried to address the current clinical need for a risk stratification tool for such patients to safely forgo surgical resection, due to the high comorbidities associated. The study included 411 treatment planning CT-scans of patients treated with neoadjuvant chemotherapy followed by surgery. The authors developed a radiomics signature that could stratify patients into low- and high-risk survival groups. The radiomic features included in the signature were found to be independent of the clinical features. Adding radiomic features to the clinical model resulted in an improvement of the predictive power (c-index) of the clinical only model from 0.67 (0.62–0.73) to 0.73 (0.66–0.80) [26]. The authors used two investigations to ensure the selection of stable radiomics features, namely test–retest and contour-recontour robustness analysis. The results signifies the added value of properly using radiomics analysis on CT scans in improving patients' risk stratification. Yet, the authors did not externally validate their signature, casting doubt on the generalizability of their signature. It is expected to be of value in cases where the scanning parameters are identical to those used in the study.

Another study by Bae et al. [27] investigated the potential of MR-based radiomics to improve the survival prediction of patients diagnosed with glioblastoma multiforme. The study is an effort to address the unmet clinical need for assessing the survival of the target group following therapy. The authors extracted radiomics features from 217 multiparametric MR scans of patients with glioblastoma. The authors identified 18 radiomics features to build a radiomic signature, and reported that the addition of radiomics features to clinical and genetic profiles of the patients significantly improves the stratification of patients [27]. The authors in this study applied a unique approach for the analysis by simultaneously analyzing radiomics features extracted from different co-registered MR sequences. The identified features were independent of the clinical and genetic factors, and the improvement in the survival prediction following their addition, supports the hypothesis of radiomics. Pitfalls in the study include the lack of assessment of radiomic feature stability before modeling, and as often seen in these studies, a lack of an external validation of the signature. However, their results support the hypothesis that radiomics are of great use when applied on scans acquired using identical settings.

Oikonomou et al. [28] reported on the potential of PET/CT-based radiomics to improve the survival stratification of patients with lung cancer treated with stereotactic body radiotherapy. The aim was to identify radiomic features that can improve the prognostication of patients following treatment. The authors extracted radiomics features from 150 PET/CT scans, and built radiomics signatures using 10 radiomics features. The authors reported that the radiomics signature was the sole predictor in the case of overall survival, and provided complementary information for the prediction of regional control [28]. The uniqueness in this study is the joint use of radiomics features extracted from the CT-component and PET-component of the PET/CT scans. The authors show how other currently used clinical parameters fail to predict overall survival, while only radiomics could. While the study highlights the potential of radiomics to improve risk stratification, no external validation of the signature was performed.

2.1.2. Progression free survival

Kirienko et al. [29] investigated the role of PET/CT-based radiomics to predict disease free survival in patients with non-small cell lung cancer undergoing surgery. The authors extracted radiomics features from PET, CT, and combined PET/CT images. The authors developed Cox regression models using only CT, only PET, and combined PET/CT radiomics features. They reported that the radiomic signatures they developed improve the current clinical stratification of the targeted patients [29]. The authors in this study investigated the reproducibility of radiomics features across the different imaging parameters in their dataset. This ensured selecting the comparable features before proceeding with signature building. The authors also provide evidence of the added value of combining radiomics features extracted from different imaging modalities. Furthermore, the ability to predict disease free survival from the time of diagnosis -which radiomics offer- improves physicians and patients decision making. However, the authors in this study did also not perform an external validation of their signature. Further validation of the signature can prompt a prospective validation trial, before incorporation into cDSS.

Another study by Kickingereder et al. [30] investigated the role of MR-based radiomics in predicting survival in patients with glioblastoma multiforme. The authors extracted radiomics features from 119 MR scans, and developed a radiomic signature using 11 features. The developed signature performed significantly better than the radiologic

and clinical risk models, and its addition to those resulted in an overall improvement of progression-free survival stratification [30]. The finding that the radiomics signature performed better than the clinical and radiologic models supports the findings reported by Bae at al. [27], and adds more evidence that radiomic features decode complementary biologic information. However, the study did not address the issues of the reproducibility and generalizability sufficiently, leaving a room for improving the performance of radiomics.

2.1.3. Tumor histology

Wu et al. [31] explored the role of radiomics in differentiating between the histologic subtypes of non-small cell lung cancer: adenocarcinoma and squamous cell carcinoma. The study was an effort to address the clinical need for less invasive and easily repeatable methods to determine tumor histology. The authors extracted radiomic features from 350 CT scans of NSCLC patients for whom the tumor histology has been determined from surgical specimens. The developed signature included 5 radiomics features, and they reported an area under the receiver characteristics curve (AUC) of 0.72 [31]. This study reflected on the potential of non-invasive radiomic signatures to differentiate between adenocarcinoma and squamous cell carcinoma. They also investigated different machine learning methodologies for building the radiomics signature. While this study generates evidence for the potential of radiomics, the performance of the developed signature is significantly lower than the current gold standard -tissue biopsy. However, there is a great room for improving the development and performance of the signature. The authors did not address the acknowledged challenges in radiomics, nor did they validate their signature on an external dataset. Preselection of reproducible features, external and prospective validation of the signature are necessary steps in the development of radiomics biomarkers.

In another study, Wu et al. [32] investigated the added value of MR-based radiomic features for the prediction of hepatocellular carcinoma (HCC) grade. The authors extracted radiomic features from 170 MRI scans of HCC patients, whose tumor grade was identified through pathological samples. The radiomics-only signature (AUC of 0.74) outperformed the clinical model (AUC of 0.60), and the combination of both significantly improved the prediction (AUC of 0.80) [32]. The authors in this study also combined radiomic features extracted from two different MR sequences and analyzed them simultaneously. The significant improvement of the predictions following the combination of clinical and radiomic features supports the independence of radiomics features from other clinical information. However, external validation of the developed signature is still a necessity before confidently performing prospective validation.

Valleries et al. [33] explored the potential of the combination of FDG-PET- and MR- based radiomics features to classify lung nodules. The authors extracted radiomics features from 51 PET and MR scans of histologically confirmed lung lesions in patients with soft-tissue sarcoma. The authors achieved a sensitivity of 0.96 and specificity of 0.93 in diagnosing metastatic nodules using a model with combined radiomic features from both PET and MR modalities. The authors used a novel interesting approach by simultaneously analyzing the features extracted from FDG-PET and MR scans, and were the first to show the potential of this method. The performance of the developed signature makes it a suitable alternative for patients for whom tissue biopsy is contraindicated. Its possible translation to cDSS might significantly improve patient outcomes, as treatment is based on the histologic diagnosis. Yet, further external and prospective validation of the signature is needed.

2.1.4. Response to therapy

Trebeschi et al. [34] explored the role of radiomics in predicting response to anti-PD1 immunotherapy in patients diagnosed with advanced melanoma and NSCLC patients. Immunotherapy has shown promising results. Yet, there is still a need for a tool to determine which patients will benefit from receiving anti-PD-1 antibodies. The authors extracted radiomic features from 1055 ROIs segmented on 203 CT scans. The authors developed a radiomic signature that could predict the response to therapy with an AUC of 0.76; showing the potential of radiomics to predict response to therapy in such patients [34]. Interestingly, the authors found correlations between the radiomic biomarker and the genes associated with cell cycle progression and mitosis. Radiomics can become a tool for assisting decision making in immunotherapy, a great unmet clinical need. The study however did not externally validate the signature, and did not sufficiently address the issues of feature stability and reproducibility. Therefore, the application of the developed signature is also limited to the patients who are scanned with the same scanning parameters as used in the training.

In a study by Horvat et al. [35], the authors investigated the role of radiomics in assessing complete clinical response (cCR) after neoadjuvant chemoradiotherapy (CRT) in patients with locally advanced rectal cancer. The guidelines of treating these patients include surgery, but evidence showed recently that a select group of patients can be safely treated with only CRT. The authors extracted radiomic features from 114 MR scans, and developed a radiomics signature with a sensitivity of 1.00, and a specificity of 0.91, which outperformed qualitative assessment of the response performed by two radiologists. The current clinical standard evaluation of cCR includes digital rectal examination and endoscopy, with an accuracy ranging between 0.71 and 0.88 [35]. The developed radiomic signature showed the highest accuracy among the available compared-with tools. Nonetheless, several steps to improve the methodology and performance of the radiomics signature could be made. The sound cCR evaluation following RCT can improve the patient management by eliminating surgical risks, time and money.

2.2. Deep learning

The application of deep learning on medical imaging could potentially fulfil more complicated tasks than hand-crafted radiomics, especially when large amounts of data are available. Furthermore, as definition of the ROIs is not a necessity in the automated deep learning workflows, the algorithm will learn patterns from the whole image and possibly make connections with the habitat of the ROIs. The applications of neural networks on medical imaging are also not limited to classification and prediction of clinical end points, but can extend to include other tasks, such as the detection and segmentation of abnormalities or target volumes, which have been investigated for decades [36]. Especially the detection and segmentation of lesions can be easily incorporated into the radiomics workflow, further automating the process. In the following paragraphs, we give examples of different applications of DL on medical imaging to perform various tasks on datasets acquired with one of the three main medical images modalities: CT, MRI, and PET.

2.2.1. Automatic segmentation of target structures

Jiang et al. [37] tried to develop a DL model that is able to accurately perform volumetric lung tumor segmentation on CT images. The authors used two versions of multiple resolution residual network models for the delineation of the ROIs. The authors used 377 tumors from the open source dataset available on The Cancer Imaging Archive (TCIA) (https://www.cancerimagingarchive.net) to train the model, and two independent datasets of 304 and 529 lung tumors to validate it. The dice similarity coefficient (DSC), which measures the spatial overlap of the segmentations, was computed to evaluate the performance of the model. The DSCs of the model on the two validation datasets were 0.75 and 0.68, respectively. The authors reported that there was no significant difference between the DL-generated mask and experts' segmentations [37]. The new approach for segmenting medical images used in this study shows to be superior to the traditional use of UNet. The approach generalizes well on external data and overcomes the multiple sizes problem. The major pitfalls is that the authors did not use the 3D geometry of the CTs to compute the results, which would probably

increase the performance significantly. The translation of such a tool to clinical practice will significantly reduce the time spent by the clinicians to plan the treatment, or evaluate the response to therapy. Moreover, from a research perspective, it can significantly reduce the time needed for radiomics research, and it will address the issue of inter-observer sensitivity of radiomics features.

In the study by Yi et al. [38], the authors developed a DL model for the segmentation of brain tumors based on 274 brain MRIs extracted from the Brain Tumor Image Segmentation Benchmark (BRATS) dataset [39]. Segmentation of brain Glioblastoma on MRI is a time-exhaustive process, and an automated, accurate and reproducible tool for this purpose is considered a clinical need. The model was trained using four different MRIs sequences. The particularity of their convolutional neural network (CNN) model is a fixed difference of Gaussian filters as a first convolution layer, as it was proven to be the most efficient for 3D segmentation. The DSC for the model was 0.89 on the BRATS dataset when compared to ground truth segmentations [38]. This article shows the superiority of 3D CNN compared to 2D CNN. The algorithm generated segmentations with a volumetric overlap of 0.89 with the experts' segmentations, which shows the potential of these tools for clinical use. However, the lack of external validation in the study limits the applicability of the algorithm to scanning parameters in the training set. The clinical practice can benefit from such tools, as it significantly reduces the time the clinicians spend, and can provide more accurate evaluation of tumor response than the current clinical routine.

Chen et al. [40] explored the possibility of developing a DL model that is able to detect and segment cervical tumors on PET imaging. The authors proposed prior information constraint CNN (PIC-CNN), which integrates a CNN with prior information of cervical tumor. The authors reported a DSC of 0.84, which was superior to the other methods tin the comparison, including transfer learning based on fully convolutional neural networks (FCN) (DSC of 0.77), automatic thresholding (DSC of 0.59), and region growing method (DSC of 0.52) [40]. The study highlights the potential of deep learning to perform well-defined and robust segmentations on PET imaging. The novelty of the approach is the use of prior information as input of the model, with delineation of the bladder. This extra information seems to give the traditional model an advantage compared to models that solely segment the tumors. However, the results were not validated on an external dataset. The application of the developed algorithm -after validating it- would decrease the need for tissue biopsy, as well as the time spent on segmenting the tumors manually or semi-automatically.

2.2.2. Oncologic classification tasks

Ardila et al. [41] tried to predict the risk of lung cancer using screening low-dose CTs. The algorithm is trained on screening low-dose CT scans of patients who were known to be at risk. The authors trained their DL model on approximately 7000 scans, and validated its performance on 1139 cases. The authors reported that the model achieved the "state-of-the-art" performance (AUC of 0.944). Furthermore, the model outperformed all the radiologists (n = 6) who were asked to give predictions. The model resulted in a significant reduction in the false positive (11%), and false negative rates (5%) [41]. While the current low-dose CT screening protocol has substantially improved in terms of consistency, it still faces major limitations represented in the inter-observer variability and incomplete characterization of image findings. The authors in [41] developed an algorithm that achieved significantly better performance than the current protocol, highlighting the potential of DL algorithms to revolutionize the field of lung cancer screening. Other advantages of the algorithm are that it eliminates the current clinical practice limitations.

Ismael et al. [42] investigated the ability of DL algorithms to classify different brain tumors. The algorithm predicts if the lesion is one of: Meningiomas, Gliomas, and Pituitary tumors. The authors developed the algorithm on 3064 T1 MRI images from 233 cancer patients. As input to the algorithm, the 2D images were considered independent from each other, and were split into 80% training and 20% testing, with strictly different patient data. The classifier used is ResNet50, a classic deep learning network, and the resultant balanced accuracy was 0.99 on a slice level and 0.97 at a patient level. This study shows that deep learning can very accurately classify brain tumors based solely on MRI data. However, the data to be used should be acquired using the same scanning parameters, as no external validation was performed in this study. There is a great clinical significance from the development of such a cDSS, as it eliminates the need for risky brain biopsies, while maintaining high accuracy.

In another study by Sibille et al. [43], the authors used the combination of CT, fluorine 18-fluorodeoxyglucose PET, atlas and PET maximum intensity projection (MIP) imaging to classify lung nodules. The study included a set of 629 patients who were diagnosed with either lung cancer or lymphoma. The authors developed models using each of imaging modalities separately, as well as in combination. The recommended algorithm achieved an AUC of 0.98 when both CT and PET were combined [43]. This study shows that the combination of CT and PET can achieve an outstanding performance in terms of predictions. The current clinical practice requires the clinician to review and classify all of the increased-uptake foci in a PET/CT scan. The algorithm could help the clinicians to quickly read those images, after highlighting the suspicious areas and their most likely classification using DL.

2.2.3. Non-oncologic classification tasks

Walsh et al. [44] explored the potential of DL to classify fibrotic lung diseases using high resolution CT scans. The current clinical guidelines for classifying fibrotic lung diseases are based on high resolution scans, and diagnoses are made based on the semantic features identified by the radiologists. While these guidelines are the current gold-standard, it suffers greatly from inter-observer variability. The authors tried to address this unmet clinical need using DL approaches. The authors trained their DL model on 929 CT scans, and tested it on 139 scans. The authors reported a performance with human-level accuracy (0.76) [44]. Of interest, the algorithm developed had a better agreement with expert radiologists than among them. The ease of application of such methods in clinical settings could benefit clinical practice, especially in centers where such clinical expertise is scarce.

In the study by Ding et al. [45], the authors tried to develop a DL model that is able to diagnose Alzheimer's disease (AD), using ^{18}F-FDG PET scans of the brain. The current clinical guidelines to diagnose AD necessitate the interpretation of scans by an expert, and there is no definitive biomarker. To investigate the potential of DL, the authors collected two datasets: one used for training and testing the model (n = 2109 scans), which was split into 90% training and 10% testing; and an independent dataset (n = 40) for the validation of the model. The authors reported an AUC of 0.98, sensitivity of 1.00 and specificity of 0.82, using scans acquired 75.8 months on average before establishing the diagnosis. The model further outperformed the readers' performance (sensitivity of 0.57 and specificity of 0.91) [45]. The significance in this study lies within the novelty of developing a biomarker for AD that is currently an unmet clinical need. In addition to the significantly better performance compared to human experts, the model can predict that the patient has AD in progression significantly earlier (~6 years). Such an application will revolutionize the clinical management of AD. However, prospective validation of this signature is needed before its translation to clinical practice.

Oh et al. [46] applied a DL based approach in order to classify the neuroimaging data related to AD. Authors used 694 MRI scans (T1-weighted MP-RAGE sequence) for solving several binary classification problems: AD vs. normal control (NC), progressive mild cognitive impairment (pMCI) vs. NC, stable mild cognitive impairment (sMCI) vs. NC and pMCI vs. sMCI. The authors utilized convolutional autoencoder-based unsupervised learning algorithms in order to classify the AD vs. NC. Following that, the authors applied a supervised transfer learning approach to classify the pMCI vs. sMCI. The developed algorithms

achieved accuracies of 0.87, 0.77, 0.63, and 0.73 for the AD, pMCI, sMCI and pMCI vs. sMCI classifications, respectively. In comparison to Ding et al. [45], the authors in this study used different DL approaches, and less numbers of patients were available for training and testing the algorithm. Furthermore, the difference in the imaging modality analysed in each study could justify the variation in performance, as AD begins with functional impairment rather than structural changes. Although the model developed by oh et al. [46] was outperformed by human experts, the authors demonstrated the possibility of end-to-end DL algorithms, which could be translated to clinical use after further optimization and prospective validation.

2.2.4. Response to therapy

Lou et al. [47] reported on the potential of DL models to predict response to radiotherapy in patients with lung cancer (primary or metastatic) using CT scans. Currently, all patients are treated similarly, while personalizing radiotherapy remains a desired, but unmet clinical need. The authors in this study collected a total of 849 scans for training the DL algorithm, and 95 scans to validate it. The authors developed a deep learning model (deep profiler) that computes and includes radiomic features in the deep-profiling process. A model combining the deep profiler and clinical variables is then used to calculate a risk score that is used to predict the response to treatment. The algorithm classifies patients into high and low risk groups, with a high performance (c-index of 0.72), which is significantly better compared to the results obtained with solely handcrafted radiomic models (c-index between 0.65 and 0.68) [47]. The algorithm developed in this study opens new potentials for individualizing radiotherapy based on patients' sensitivity. Thereby, avoiding over- or under-treatment, and the side-effects of unnecessary treatment. Nevertheless, proper prospective validation of the developed algorithm remains a necessity.

Ypsilantis et al. [48] used convolutional neural networks to develop a model that is capable of predicting response to neo-adjuvant chemotherapy (NAC) in patients with esophageal cancer using PET scans. NAC is considered a standard of care in some cancers. While NAC has favourable outcomes in patients who respond, patients who do not end up with worse outcomes. To investigate the potential of QIA techniques, the authors collected 107 PET scans of patients diagnosed with esophageal cancer, treated with NAC, and followed-up to determine response. The authors compared the performance of hand-crafted radiomics with deep learning approaches. The authors reported that the developed deep learning algorithm outperformed the hand-crafted radiomics model, and achieved a sensitivity of 0.81 and specificity of 0.82 [48]. The algorithm developed in this study highlights the potential of using DL to predict patients' response to therapy at baseline, which is considered a substantial clinical added value.

3. Challenges and future directions

Biomarkers are defined as "objective indications of medical state observed from outside the patient – which can be measured accurately and reproducibly" [49]. The core of choosing a biomarker is the ability to measure it objectively. The reproducibility of imaging quantitative features across different imaging parameters is currently the steepest hurdle in QIA. As more research is being performed, other challenges, such as the sensitivity of QIA features to variations in the segmentation of the ROIs; and the lack of feature reproducibility across different implementations of radiomics toolboxes, are becoming increasingly clear.

3.1. The stability and reproducibility of quantitative features

Since the first landmark study in radiomics by Aerts et al. [50], the sensitivity of radiomic features to repeated acquisitions has been acknowledged. The authors performed a test-retest stability investigation, and used 100 out of 440 calculated radiomic features based on the stability rank of the features. The authors also acknowledged the sensitivity of features to differences in segmentations, and performed a primary feature selection based on the features' robustness with regards to differences in both test-retest and segmentations. More recently, several studies reported on the sensitivity of radiomic features to temporal changes in test-retest studies across different modalities, including CT, MRI, and PET.

3.1.1. Anatomical imaging

Anatomical imaging (CT and MRI) is used to explore the underlying anatomical structures. CT imaging is standardized using the hounsfield units (HU) [51]. On the other hand, MR imaging has no such standardized intensity measurements [52]. Even though CT imaging uses standardized measurements, CT-based radiomics are not necessarily reproducible. Several studies reported that a significant number of CT-based radiomic features are not reproducible in test-retest settings, where the scans are acquired using the same scanning parameters [53–55]. Other studies that investigated the reproducibility of CT-based radiomics features across different imaging acquisition and reconstruction parameters reported that the majority of radiomic features are significantly affected by such differences [53,56,57]. Unreproducible radiomic features should be removed before starting the modeling of radiomics signatures. Therefore, it is always necessary to perform preselection of stable radiomics features based on the data under study, before starting the modeling.

MR-based radiomics is even more complex and challenging to standardize compared to CT based radiomics, as more factors -in addition to lack of standardized intensity measurements- affect MR imaging [58]. Some studies reported on the stability of various MR-based features. Fiset et al. [59] investigated the reproducibility of T2-weighted MRI of cervical cancer in three different settings: (i) test–retest; (ii) diagnostic MRI versus simulation MRI; (iii) interobserver variability. The authors reported that 22.6%, 6.2% and 74.4% of 1761 extracted radiomics features were reproducible across test-retest, diagnostic versus simulation MRI, and different observers, respectively. Semi-parametric maps derived from specialized MRI sequences suffer less from the lack of stability: Peerlings et al. [60] reported on the stability of radiomics features extracted from apparent diffusion coefficient (ADC) map in test-retest and across different cancer types, centers, and vendors. The authors reported that out of 1322 extracted radiomics features, 122 features were stable across all cancers, centers, and vendors.

On top of these challenges, using contrast agents for imaging adds another level of complexity to the reproducibility of features, due to the differences in the cardiac function of patients being scanned. Changes in cardiac function can affect the time the distribution of the contrast in the body takes [61]. Another factor in contrast-enhanced images is the difference in time between the injection of the contrast and scan acquisition, which might be slightly different across centers and protocols.

3.1.2. Functional imaging

Functional imaging is used to assess the metabolic activity of a region of interest, and includes the injection of radiopharmaceuticals. Some standardized measurements in PET are already being extracted and used in clinical practice, such as the standardized uptake value (SUV), and the metabolically active tumor volume (MTV) [7].

The challenges of radiomics for functional imaging are similar to the challenges of contrast-enhanced anatomical imaging radiomics, where the variability in the injected radiopharmaceutical activity, the time between injection and image acquisition, and acquisition time per bed position have profound implications on the reproducibility of radiomics features [62]. In addition, functional imaging lacks anatomical specificity and suffers from low resolution, which could be addressed by the use of hybrid imaging [22]. Tixier et al. [63] investigated the reproducibility of SUV measurements, intensity histogram features, intensity-size zone features, and co-occurrence matrices features. The authors

acquired two ^{18}F-FDG PET scans of 16 patients, with a 4-days' time interval. In contrast to further studies, the authors reported that these features were insensitive to the discretization range. Hatt et al. [64] investigated the robustness of PET based heterogeneity textural features with respect to the delineation of functional volumes and partial volume effects correction. The authors reported that these features were significantly affected by the differences in the delineation. The authors further reported that local features, e.g entropy and heterogeneity, were more robust when compared to regional features, e.g intensity variability and size-zone variability. Leijenaar et al. [65] investigated the role of SUV discretization on radiomics features. The authors used two different methods for SUV discretization, and reported that differences in SUV discretization significantly affect the reproducibility of ^{18}F-FDG PET based radiomic features. The authors recommended the standardization of methodology for radiomics analysis. Altazi et al. [66] investigated the reproducibility of PET based radiomic features in cervical cancer patients. The authors investigated the reproducibility in three different scenarios: (i) manual versus computer-aided segmentations, (ii) gray-level discretization, and (iii) reconstruction algorithms. The authors extracted 79 PET radiomics features, and reported that the percentage of stable features in the three scenarios were 13%, 5%, and 1% respectively. Shiri et al. [67] explored the effects of different reconstruction on ^{18}F-FDG PET radiomics. The authors studied the effects of several factors including number of sub-iterations, number of subsets, full width at half maximum (FWHM) of Gaussian filter, and scan time per bed position and matrix size. The authors reported that 47% of the features were found to be robust, and these include shape, 44% of the intensity based features, and 41% of the texture based features. However, with changes in matrix size, the authors reported that only 6% of the features were robust.

The discrepancies in the reported percentages of stable/reproducible features across the reported studies are most likely linked to the variations between the datasets used in each of the studies in the scanners, and scans acquisition and reconstruction parameters combinations. However, these discrepancies are expected because of the different complexity of radiomics features, as well as the interaction between the different scanning parameters. All of the above mentioned studies reported that a variable percentage of radiomics features are affected, which highlights the necessity of performing feature stability/reproducibility studies based on the data under analysis before performing radiomics analysis.

3.2. Sensitivity of quantitative imaging features to variations in the segmentation of the ROIs

In QIA, the medical images are converted to numerical arrays before feature calculation. Consequently, it is intuitive that differences in segmentations affect the quantitative imaging feature values variably, depending on the feature definition. Many studies have identified lists of radiomics features that are robust to variability in segmentations [50,68,69]. Furthermore, with the inclusion of deep learning methods in image analysis, efforts are being made to develop reliable and reproducible automatic segmentations of various regions of interest as described in 3.2.1. Deep learning suffers less in this aspect, as the provision of ROIs is not obligatory.

3.3. The different implementations of radiomics feature extraction toolboxes

It is common knowledge in the radiomics community that different radiomics toolboxes use different pre-processing techniques and/or feature definitions, which lead(s) to variations in estimation of radiomics feature values when different software solutions are used. To address this issue, the radiomics community started an initiative – Imaging Biomarkers Standardization Initiative (IBSI) – that aims at standardizing radiomics feature extraction using different toolboxes [70]. To date, the IBSI standardized the extraction of 169 radiomics features [71]. Limiting the radiomics analysis to the IBSI standardized features can facilitate radiomic features interchangeability across platforms.

3.4. Future directions

To address the issue of radiomic features reproducibility, some harmonization methods have been investigated in the literature. Of the trending methods is Combine Batches (ComBat). ComBat is a statistical method that was developed to remove the batch effects in microarray expressions [72]. While several studies have reported on the application of ComBat harmonization in radiomics analysis as a means to remove batch effects [73,74], its direct application on radiomics data is not in concordance with the mathematical definition of ComBat [72], or with the hypothesis that radiomics correlate with biology. This is because ComBat assumes that the differences between batches are attributed to two groups of factors, the first group refers to the biological covariates, which radiomics features are investigated for correlations with. Moreover, adding biologic covariates for ComBat in the training of radiomics signatures will hinder its prospective use, because it will be the outcome the radiomic signature tries to predict. The second group refers to the "non-biologic" factors, such as image acquisition and reconstruction parameters. ComBat was defined to handle one batch effect at a time. In contrast to gene expression arrays for which ComBat was designed, radiomic features have different complexity levels, which are expected to be non-uniformly affected by the variations in imaging parameters. In addition, the differences in image acquisition and reconstruction settings in a given retrospective imaging dataset are usually in more than one imaging parameter. The proper use of ComBat would require the assessment of the reproducibility of radiomics features after applying ComBat on representative objects with no biologic variations, such as phantoms. Then, radiomics features extracted from patients' scans acquired with the same imaging parameters can be transformed based on the location/scale parameters estimated by the application of ComBat on the phantom data. We here propose a framework for performing robust radiomics analysis (Fig. 3). Nonetheless, a radiomics-specific harmonization method is still needed to eliminate the need for phantom studies, as the performance of ComBat is expected to be dependent on the variations in scanning parameters in the data.

The workflow consists of consecutive steps, and can be used to preselect reproducible and harmonizable radiomics features. The first step in the workflow is the collection of retrospective patient imaging data to be analyzed. In the second step, scan acquisition and reconstruction parameters must be extracted from the collected patient data. The next step includes scanning a phantom with the parameters extracted from the patient imaging data. This allows the assessment of the reproducibility of radiomics features across the different scan acquisition and reconstruction parameters, and the selection of those features for performing robust radiomics analysis.

Based on our review of existing literature and our own experience, in order to use ComBat in the context of radiomics analysis (steps 5–7), two extra steps are needed. After selecting the features that are insensitive to the variations in the scanning parameters extracted from the patient data, features that are reproducible in test-retest in each of the combinations of those scanning parameters must be identified. ComBat is then applied on the features that are reproducible in test-retest but not across different scanning parameters. The concordance of Radiomic features is assessed following the application of ComBat. The location/scale shift parameters estimated by performing ComBat on the phantom data are then applied to the radiomics features extracted from patient data to harmonize them. The combination of the identified stable and harmonizable features can be further used to build the radiomics signature.

The challenges discussed above raise questions about the future applications of radiomics, and the development of radiomic signatures as clinical biomarkers. To begin with, how to approach the concept of external validation in radiomics studies. Do radiomic signatures need to

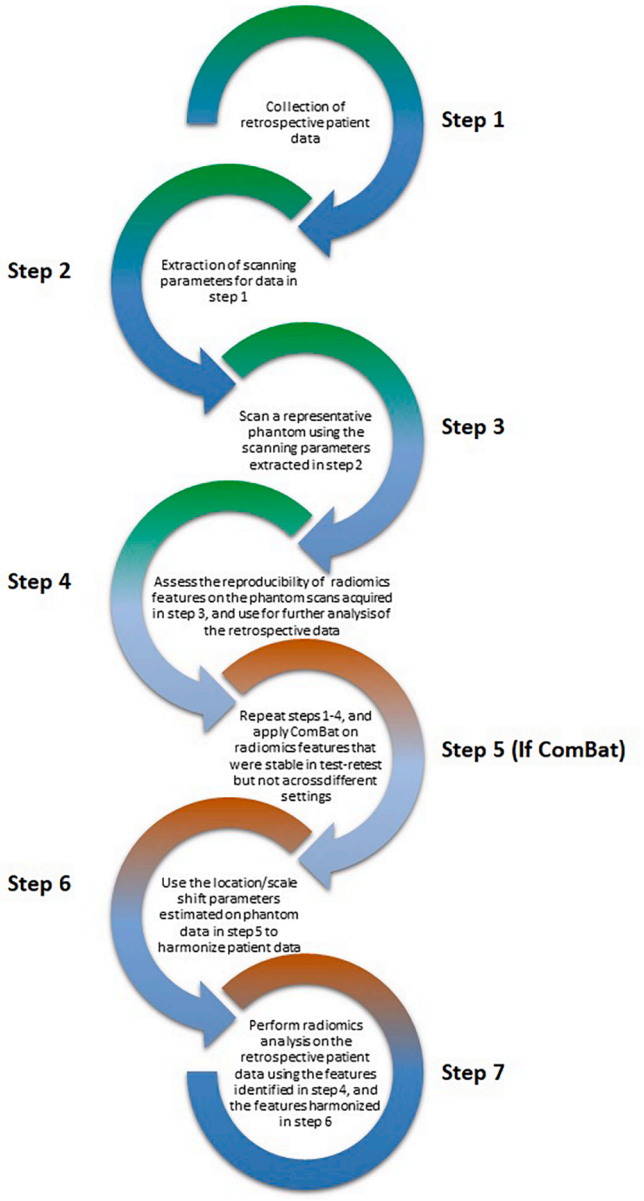

Fig. 3. Proposed workflow for robust radiomics analysis.

be externally validated as is the case with other biomarkers, given all the challenges of reproducibility across different imaging settings? Or would the observatory prospective validation of a given signature in a specific image setting suffice? Does the development of radiomic signatures need to be specific for a scanner model and imaging settings? The ultimate solution will be the development of specific quantitative imaging parameters, as there is currently a clinical direction to personalize imaging settings per patient, which will have its toll on radiomics analysis. The direct application of radiomics analysis on data acquired heterogeneously could lead to spurious results, and inability of translating the results in a meaningful manner.

4. Conclusion

Quantitative imaging techniques (radiomics and deep learning) present a perfect candidate for personalizing patients' management. Applying these techniques in a sound manner can provide highly accurate and reproducible tools that minimize costs and time loss. However, to incorporate QIA in cDSS, the quantitative features should fulfil the definition of a biomarker, namely the stability and reproducibility. The future of quantitative image analysis in general lies within harmonizing the imaging protocols across centers and scanners, or within the development of a unique global protocol for quantitative analysis scans. Hence, the development of radiomics-specific tools to harmonize medical images and facilitate meaningful quantitative image analysis of the currently available retrospective data remains a necessity. Our proposed framework is expected to improve the robustness of radiomics analysis. Nevertheless, the benefits of the proper application and translation of QIA on medical imaging are undoubted. QIA techniques will be a valuable asset for both: the clinicians and patients. QIA can become an efficient means for aiding clinicians in risk stratification, early diagnosis, and improved management of patients.

Competing interests

Dr. Philippe Lambin reports, within and outside the submitted work, grants/sponsored research agreements from Varian medical, Oncoradiomics, ptTheragnostic, Health Innovation Ventures and DualTpharma. He received an advisor/presenter fee and/or reimbursement of travel costs/external grant writing fee and/or in kind manpower contribution from Oncoradiomics, BHV, Merck and Convert pharmaceuticals. Dr. Lambin has shares in the company Oncoradiomics SA and Convert pharmaceuticals SA and is co-inventor of two issued patents with royalties on radiomics (PCT/NL2014/050248, PCT/NL2014/050728) licensed to Oncoradiomics and one issue patent on mtDNA (PCT/EP2014/059089) licensed to ptTheragnostic/DNAmito, three non-patentable invention (softwares) licensed to ptTheragnostic/DNAmito, Oncoradiomics and Health Innovation Ventures. Dr. Woodruff has (minority) shares in the company Oncoradiomics.

CRediT authorship contribution statement

A. Ibrahim: Conceptualization, Methodology, Formal analysis, Data curation, Writing - original draft, Project administration. **S. Primakov:** Formal analysis, Data curation, Writing - original draft, Visualization. **M. Beuque:** Formal analysis, Data curation, Writing - original draft. **H. C. Woodruff:** Supervision, Writing - review & editing. **I. Halilaj:** Visualization. **G. Wu:** Resources, Data curation. **T. Refaee:** Resources. **R. Granzier:** Resources. **Y. Widaatalla:** Resources. **R. Hustinx:** Supervision. **F.M. Mottaghy:** Supervision, Writing - review & editing. **P. Lambin:** Conceptualization, Methodology, Writing - review & editing, Project administration, Supervision.

Acknowledgments

Authors acknowledge financial support from ERC advanced grant (ERC-ADG-2015, n° 694812 - Hypoximmuno), ERC-2018-PoC (n° 81320 – CL-IO). We further acknowledge the financial support from Maastricht-Liege imaging valley grant. This research is also supported by the Dutch technology Foundation STW (grant n° P14-19 Radiomics STRaTegy), which is the applied science division of NWO, and the Technology Programme of the Ministry of Economic Affairs. Authors also acknowledge financial support from SME Phase 2 (RAIL – n°673780), EUROSTARS (DART, DECIDE), the European Program H2020-2015-17 (BD2Decide - PHC30-689715, ImmunoSABR – n° 733008, PREDICT – ITN – n° 766276), TRANSCAN Joint Transnational Call 2016 (JTC2016 'CLEARLY'- n° UM 2017-8295) and Interreg V-A Euregio Meuse-Rhine ('Euradiomics'). Authors further acknowledge financial support by the Dutch Cancer Society (KWF Kankerbestrijding), Project number 12085/2018-2.

References

[1] S. Walsh, E.E.C. de Jong, J.E. van Timmeren, A. Ibrahim, I. Compter, J. Peerlings, S. Sanduleanu, T. Refaee, S. Keek, R.T.H.M. Larue, Y. van Wijk, A.J.G. Even,

A. Jochems, M.S. Barakat, R.T.H. Leijenaar, P. Lambin, Decision support systems in oncology, JCO Clin. Cancer Inform. 3 (2019) 1–9, https://doi.org/10.1200/CCI.18.00001.

[2] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G.P.M. van Stiphout, P. Granton, C.M.L. Zegers, R. Gillies, R. Boellard, A. Dekker, H.J.W.L. Aerts, Radiomics: extracting more information from medical images using advanced feature analysis, Eur. J. Cancer 48 (2012) 441–446, https://doi.org/10.1016/j.ejca.2011.11.036.

[3] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, Radiology 278 (2016) 563–577, https://doi.org/10.1148/radiol.2015151169.

[4] P. Lambin, R.T.H. Leijenaar, T.M. Deist, J. Peerlings, E.E.C. de Jong, J. van Timmeren, S. Sanduleanu, R.T.H.M. Larue, A.J.G. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T. Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F. M. Mottaghy, J.E. Wildberger, S. Walsh, Radiomics: the bridge between medical imaging and personalized medicine, Nat. Rev. Clin. Oncol. 14 (2017) 749–762, https://doi.org/10.1038/nrclinonc.2017.141.

[5] T. Refaee, G. Wu, A. Ibrahim, I. Halilaj, R.T.H. Leijenaar, W. Rogers, H.A. Gietema, L.E.L. Hendriks, P. Lambin, H.C. Woodruff, The emerging role of radiomics in COPD and lung cancer, Respiration 99 (2020) 99–107, https://doi.org/10.1159/000505429.

[6] R.C. Hardie, S.K. Rogers, T. Wilson, A. Rogers, Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs, Med. Image Anal. 12 (2008) 240–258, https://doi.org/10.1016/j.media.2007.10.004.

[7] G.J.R. Cook, M. Siddique, B.P. Taylor, C. Yip, S. Chicklore, V. Goh, Radiomics in PET: principles and applications, Clin. Transl. Imaging 2 (2014) 269–276, https://doi.org/10.1007/s40336-014-0064-0.

[8] S. Chicklore, V. Goh, M. Siddique, A. Roy, P.K. Marsden, G.J.R. Cook, Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis, Eur. J. Nucl. Med. Mol. Imaging 40 (2013) 133–140, https://doi.org/10.1007/s00259-012-2247-0.

[9] C. Swanton, Intratumor heterogeneity: evolution through space and time, Cancer Res. 72 (2012) 4875–4882, https://doi.org/10.1158/0008-5472.CAN-12-2217.

[10] M. Gerlinger, A.J. Rowan, S. Horswell, M. Math, J. Larkin, D. Endesfelder, E. Gronroos, P. Martinez, N. Matthews, A. Stewart, P. Tarpey, I. Varela, B. Phillimore, S. Begum, N.Q. McDonald, A. Butler, D. Jones, K. Raine, C. Latimer, C.R. Santos, M. Nohadani, A.C. Eklund, B. Spencer-Dene, G. Clark, L. Pickering, G. Stamp, M. Gore, Z. Szallasi, J. Downward, P.A. Futreal, C. Swanton, Intratumor heterogeneity and branched evolution revealed by multiregion sequencing, N. Engl. J. Med. 366 (2012) 883–892, https://doi.org/10.1056/NEJMoa1113205.

[11] T.M. Soo, M. Bernstein, J. Provias, R. Tasker, A. Lozano, A. Guha, Failed stereotactic biopsy in a series of 518 cases, Stereotact. Funct. Neurosurg. 64 (1995) 183–196, https://doi.org/10.1159/000098747.

[12] S.S.F. Yip, H.J.W.L. Aerts, Applications and limitations of radiomics, Phys. Med. Biol. 61 (2016) R150–66, https://doi.org/10.1088/0031-9155/61/13/R150.

[13] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444, https://doi.org/10.1038/nature14539.

[14] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. 5 (1943) 115–133, https://doi.org/10.1007/BF02478259.

[15] L. Hongtao, Z. Qinchuan, Applications of Deep Convolutional Neural Network in Computer Vision, J. Data Acquisition Process. (2016). http://en.cnki.com.cn/Article_en/CJFDTotal-SJCJ201601001.htm.

[16] H. Shirani-Mehr, Applications of deep learning to sentiment analysis of movie reviews, Tech. Rep. NAVTRADEVCEN (2014) 1–8.

[17] L. Deng, A tutorial survey of architectures, algorithms, and applications for deep learning, APSIPA Trans. Signal Information Processing 3 (2014), https://doi.org/10.1017/atsip.2013.9.

[18] K. Janocha, W.M. Czarnecki, On loss functions for deep neural networks in classification, Schedae Informaticae. 1/2016 (2017). doi: 10.4467/20838476si.16.004.6185.

[19] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, Deep learning for computer vision: a brief review, Comput. Intell. Neurosci. 2018 (2018) 7068349, https://doi.org/10.1155/2018/7068349.

[20] R. Simhambhatla, K. Okiah, S. Kuchkula, R. Slater, Self-Driving Cars: Evaluation of Deep Learning Techniques for Object Detection in Different Driving Conditions, SMU Data Science Review. 2 (2019) 23. https://scholar.smu.edu/datasciencereview/vol2/iss1/23/ (accessed May 14, 2020).

[21] D. Shen, G. Wu, H.-I. Suk, Deep learning in medical image analysis, Annu. Rev. Biomed. Eng. 19 (2017) 221–248, https://doi.org/10.1146/annurev-bioeng-071516-044442.

[22] A. Ibrahim, M. Vallières, H. Woodruff, S. Primakov, M. Beheshti, S. Keek, T. Refaee, S. Sanduleanu, S. Walsh, O. Morin, P. Lambin, R. Hustinx, F.M. Mottaghy, Radiomics analysis for clinical decision support in nuclear medicine, Semin. Nucl. Med. 49 (2019) 438–449, https://doi.org/10.1053/j.semnuclmed.2019.06.005.

[23] L.R. Cardon, H. Watkins, Waiting for the working draft from the human genome project. A huge achievement, but not of immediate medical use, BMJ 320 (2000) 1223–1224, https://doi.org/10.1136/bmj.320.7244.1223.

[24] N.J. Schork, Personalized medicine: time for one-person trials, Nature 520 (2015) 609–611, https://doi.org/10.1038/520609a.

[25] C. Parmar, J.D. Barry, A. Hosny, J. Quackenbush, H.J.W.L. Aerts, Data analysis strategies in medical imaging, Clin. Cancer Res. 24 (2018) 3492–3499, https://doi.org/10.1158/1078-0432.CCR-18-0385.

[26] J. Wang, L. Shen, H. Zhong, Z. Zhou, P. Hu, J. Gan, R. Luo, W. Hu, Z. Zhang, Radiomics features on radiotherapy treatment planning CT can predict patient survival in locally advanced rectal cancer patients, Sci. Rep. 9 (2019) 15346, https://doi.org/10.1038/s41598-019-51629-4.

[27] S. Bae, Y.S. Choi, S.S. Ahn, J.H. Chang, S.-G. Kang, E.H. Kim, S.H. Kim, S.-K. Lee, Radiomic MRI Phenotyping of Glioblastoma: Improving Survival Prediction, Radiology 289 (2018) 797–806, https://doi.org/10.1148/radiol.2018180200.

[28] A. Oikonomou, F. Khalvati, P.N. Tyrrell, M.A. Haider, U. Tarique, L. Jimenez-Juan, M.C. Tjong, I. Poon, A. Eilaghi, L. Ehrlich, P. Cheung, Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy, Sci. Rep. 8 (2018) 4003, https://doi.org/10.1038/s41598-018-22357-y.

[29] M. Kirienko, L. Cozzi, L. Antunovic, L. Lozza, A. Fogliata, E. Voulaz, A. Rossi, A. Chiti, M. Sollini, Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery, Eur. J. Nucl. Med. Mol. Imaging 45 (2018) 207–217, https://doi.org/10.1007/s00259-017-3837-7.

[30] P. Kickingereder, S. Burth, A. Wick, M. Götz, O. Eidel, H.-P. Schlemmer, K. H. Maier-Hein, W. Wick, M. Bendszus, A. Radbruch, D. Bonekamp, Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models, Radiology 280 (2016) 880–889, https://doi.org/10.1148/radiol.2016160845.

[31] W. Wu, C. Parmar, P. Grossmann, J. Quackenbush, P. Lambin, J. Bussink, R. Mak, H.J.W.L. Aerts, Exploratory study to identify radiomics classifiers for lung cancer histology, Front. Oncol. 6 (2016) 71, https://doi.org/10.3389/fonc.2016.00071.

[32] M. Wu, H. Tan, F. Gao, J. Hai, P. Ning, J. Chen, S. Zhu, M. Wang, S. Dou, D. Shi, Predicting the grade of hepatocellular carcinoma based on non-contrast-enhanced MRI radiomics signature, Eur. Radiol. 29 (2019) 2802–2811, https://doi.org/10.1007/s00330-018-5787-2.

[33] M. Vallières, C.R. Freeman, S.R. Skamene, I. El Naqa, A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities, Phys. Med. Biol. 60 (2015) 5471–5496, https://doi.org/10.1088/0031-9155/60/14/5471.

[34] S. Trebeschi, S.G. Drago, N.J. Birkbak, I. Kurilova, A.M. Calin, A. Delli Pizzi, F. Lalezari, D.M.J. Lambregts, M. W. Rohaan, C. Parmar, E.A. Rozeman, K.J. Hartemink, C. Swanton, J.B.A.G. Haanen, C.U. Blank, E.F. Smit, R.G.H. Beets-Tan, H.J.W.L. Aerts, Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers, Annals of. (2019). https://academic.oup.com/annonc/article-abstract/30/6/998/5416144.

[35] N. Horvat, H. Veeraraghavan, M. Khan, I. Blazic, J. Zheng, M. Capanu, E. Sala, J. Garcia-Aguilar, M.J. Gollub, I. Petkovska, MR imaging of rectal cancer: radiomics analysis to assess treatment response after neoadjuvant therapy, Radiology 287 (2018) 833–843, https://doi.org/10.1148/radiol.2018172300.

[36] J. Alirezaie, M.E. Jernigan, C. Nahmias, Automatic segmentation of cerebral MR images using artificial neural networks, IEEE Trans. Nucl. Sci. 45 (1998) 2174–2182, https://doi.org/10.1109/23.708336.

[37] J. Jiang, Y.-C. Hu, C.-J. Liu, D. Halpenny, M.D. Hellmann, J.O. Deasy, G. Mageras, H. Veeraraghavan, Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images, IEEE Trans. Med. Imaging 38 (2019) 134–144, https://doi.org/10.1109/TMI.2018.2857800.

[38] D. Yi, M. Zhou, Z. Chen, O. Gevaert, 3-D Convolutional Neural Networks for Glioblastoma Segmentation, arXiv [cs.CV]. (2016). http://arxiv.org/abs/1611.04534.

[39] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B.B. Avants, N. Ayache, P. Buendia, D.L. Collins, N. Cordier, J.J. Corso, A. Criminisi, T. Das, H. Delingette, Ç. Demiralp, C.R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N.M. John, E. Konukoglu, D. Lashkari, J.A. Mariz, R. Meier, S. Pereira, D. Precup, S.J. Price, T.R. Raviv, S.M.S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H.-C. Shin, J. Shotton, C.A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T.J. Taylor, O.M. Thomas, N.J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D.H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, K. Van Leemput, The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Trans. Med. Imaging 34 (2015) 1993–2024, https://doi.org/10.1109/TMI.2014.2377694.

[40] L. Chen, C. Shen, S. Li, G. Maquilan, K. Albuquerque, M.R. Folkert, J. Wang, Automatic PET cervical tumor segmentation by deep learning with prior information, in: Medical Imaging 2018: Image Processing, International Society for Optics and Photonics, 2018: p. 1057436. doi: 10.1117/12.2293926.

[41] D. Ardila, A.P. Kiraly, S. Bharadwaj, B. Choi, J.J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, D.P. Naidich, S. Shetty, Author Correction: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, Nat. Med. 25 (2019) 1319, https://doi.org/10.1038/s41591-019-0536-x.

[42] S.A. Abdelaziz Ismael, A. Mohammed, H. Hefny, An enhanced deep learning approach for brain cancer MRI images classification using residual networks, Artif. Intell. Med. 102 (2020), 101779, https://doi.org/10.1016/j.artmed.2019.101779.

[43] L. Sibille, R. Seifert, N. Avramovic, T. Vehren, B. Spottiswoode, S. Zuehlsdorff, M. Schäfers, 18F-FDG PET/CT uptake classification in lymphoma and lung cancer by using deep convolutional neural networks, Radiology 294 (2020) 445–452, https://doi.org/10.1148/radiol.2019191114.

[44] S.L.F. Walsh, L. Calandriello, M. Silva, N. Sverzellati, Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study, Lancet Respir. Med. 6 (2018) 837–845, https://doi.org/10.1016/S2213-2600(18)30286-8.

[45] Y. Ding, J.H. Sohn, M.G. Kawczynski, H. Trivedi, R. Harnish, N.W. Jenkins, D. Lituiev, T.P. Copeland, M.S. Aboian, C. Mari Aparici, S.C. Behr, R.R. Flavell, S.-

Y. Huang, K.A. Zalocusky, L. Nardo, Y. Seo, R.A. Hawkins, M. Hernandez Pampaloni, D. Hadley, B.L. Franc, A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain, Radiology 290 (2019) 456–464, https://doi.org/10.1148/radiol.2018180958.

[46] K. Oh, Y.-C. Chung, K.W. Kim, W.-S. Kim, I.-S. Oh, Author Correction: Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning, Sci. Rep. 10 (2020) 5663, https://doi.org/10.1038/s41598-020-62490-1.

[47] B. Lou, S. Doken, T. Zhuang, D. Wingerter, M. Gidwani, N. Mistry, L. Ladic, A. Kamen, M.E. Abazeed, An image-based deep learning framework for individualising radiotherapy dose: a retrospective analysis of outcome prediction, The Lancet Digital Health 1 (2019) e136–e147, https://doi.org/10.1016/s2589-7500(19)30058-5.

[48] P.-P. Ypsilantis, M. Siddique, H.-M. Sohn, A. Davies, G. Cook, V. Goh, G. Montana, Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks, PLoS ONE 10 (2015), e0137036, https://doi.org/10.1371/journal.pone.0137036.

[49] K. Strimbu, J.A. Tavel, What are biomarkers? Curr. Opin. HIV AIDS 5 (2010) 463. https://www.ncbi.nlm.nih.gov/pmc/articles/pmc3078627/.

[50] H.J.W.L. Aerts, E.R. Velazquez, R.T.H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M.M. Rietbergen, C.R. Leemans, A. Dekker, J. Quackenbush, R.J. Gillies, P. Lambin, Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nat. Commun. 5 (2014) 4006, https://doi.org/10.1038/ncomms5006.

[51] U. Schneider, E. Pedroni, A. Lomax, The calibration of CT Hounsfield units for radiotherapy treatment planning, Phys. Med. Biol. 41 (1996) 111–124, https://doi.org/10.1088/0031-9155/41/1/009.

[52] L.G. Nyúl, J.K. Udupa, On standardizing the MR image intensity scale, Magn. Reson. Med. 42 (1999) 1072–1081. https://doi.org/3.0.co;2-m.">10.1002/(sici)1522-2594(199912)42:6<1072::aid-mrm11>3.0.co;2-m.

[53] R. Berenguer, M.D.R. Pastor-Juan, J. Canales-Vázquez, M. Castro-García, M.V. Villas, F. Mansilla Legorburo, S. Sabater, Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters, Radiology. 288 (2018) 407–415. doi: 10.1148/radiol.2018172361.

[54] J.E. van Timmeren, R.T.H. Leijenaar, W. van Elmpt, J. Wang, Z. Zhang, A. Dekker, P. Lambin, Test-retest data for radiomics feature stability analysis: generalizable or study-specific? Tomography. 2 (2016) 361–365, https://doi.org/10.18383/j.tom.2016.00208.

[55] L. Lu, R.C. Ehmke, L.H. Schwartz, B. Zhao, Assessing agreement between radiomic features computed for multiple CT imaging settings, PLoS ONE 11 (2016), e0166550, https://doi.org/10.1371/journal.pone.0166550.

[56] D. Mackin, X. Fave, L. Zhang, D. Fried, J. Yang, B. Taylor, E. Rodriguez-Rivera, C. Dodge, A.K. Jones, L. Court, Measuring computed tomography scanner variability of radiomics features, Invest. Radiol. 50 (2015) 757–765, https://doi.org/10.1097/RLI.0000000000000180.

[57] I. Zhovannik, J. Bussink, A. Traverso, Z. Shi, P. Kalendralis, L. Wee, A. Dekker, R. Fijten, R. Monshouwer, Learning from scanners: bias reduction and feature correction in radiomics, Clin. Transl. Radiat. Oncol. 19 (2019) 33–38, https://doi.org/10.1016/j.ctro.2019.07.003.

[58] A. Webb, G.C. Kagadis, Introduction to Biomedical Imaging, Med. Phys. 30 (2003) 2267–2267. doi: 10.1118/1.1589017.

[59] S. Fiset, M.L. Welch, J. Weiss, M. Pintilie, J.L. Conway, M. Milosevic, A. Fyles, A. Traverso, D. Jaffray, U. Metser, J. Xie, K. Han, Repeatability and reproducibility of MRI-based radiomic features in cervical cancer, Radiother. Oncol. 135 (2019) 107–114, https://doi.org/10.1016/j.radonc.2019.03.001.

[60] J. Peerlings, H.C. Woodruff, J.M. Winfield, A. Ibrahim, B.E. Van Beers, A. Heerschap, A. Jackson, J.E. Wildberger, F.M. Mottaghy, N.M. DeSouza, P. Lambin, Stability of radiomics features in apparent diffusion coefficient maps from a multi-centre test-retest trial, Sci. Rep. 9 (2019) 4800, https://doi.org/10.1038/s41598-019-41344-5.

[61] K.T. Bae, Intravenous contrast medium administration and scan timing at CT: considerations and approaches, Radiology 256 (2010) 32–61, https://doi.org/10.1148/radiol.10090908.

[62] G.J.R. Cook, G. Azad, K. Owczarczyk, M. Siddique, V. Goh, Challenges and promises of PET radiomics, Int. J. Radiat. Oncol. Biol. Phys. 102 (2018) 1083–1089, https://doi.org/10.1016/j.ijrobp.2017.12.268.

[63] F. Tixier, M. Hatt, C.C. Le Rest, A. Le Pogam, L. Corcos, D. Visvikis, Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET, J. Nucl. Med. 53 (2012) 693–700, https://doi.org/10.2967/jnumed.111.099127.

[64] M. Hatt, F. Tixier, C.C. Le Rest, O. Pradier, Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma, European Journal of. (2013). https://link.springer.com/article/10.1007/s00259-013-2486-8.

[65] R.T.H. Leijenaar, G. Nalbantov, S. Carvalho, W.J.C. van Elmpt, E.G.C. Troost, R. Boellaard, H.J.W.L. Aerts, R.J. Gillies, P. Lambin, The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis, Sci. Rep. 5 (2015) 11075, https://doi.org/10.1038/srep11075.

[66] B.A. Altazi, G.G. Zhang, D.C. Fernandez, M.E. Montejo, D. Hunt, J. Werner, M. C. Biagioli, E.G. Moros, Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms, J. Appl. Clin. Med. Phys. 18 (2017) 32–48. https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/acm2.12170.

[67] I. Shiri, A. Rahmim, P. Ghaffarian, P. Geramifar, H. Abdollahi, A. Bitarafan-Rajabi, The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies, Eur. Radiol. 27 (2017) 4498–4509, https://doi.org/10.1007/s00330-017-4859-z.

[68] R.T.H. Leijenaar, S. Carvalho, E.R. Velazquez, W.J.C. van Elmpt, C. Parmar, O. S. Hoekstra, C.J. Hoekstra, R. Boellaard, A.L.A.J. Dekker, R.J. Gillies, H.J.W. L. Aerts, P. Lambin, Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability, Acta Oncol. 52 (2013) 1391–1397, https://doi.org/10.3109/0284186X.2013.812798.

[69] M. Pavic, M. Bogowicz, X. Würms, S. Glatz, T. Finazzi, O. Riesterer, J. Roesch, L. Rudofsky, M. Friess, P. Veit-Haibach, M. Huellner, I. Opitz, W. Weder, T. Frauenfelder, M. Guckenberger, S. Tanadini-Lang, Influence of inter-observer delineation variability on radiomics stability in different tumor sites, Acta Oncol. 57 (2018) 1070–1074, https://doi.org/10.1080/0284186X.2018.1445283.

[70] M. Hatt, M. Vallieres, D. Visvikis, A. Zwanenburg, IBSI: an international community radiomics standardization initiative, J. Nucl. Med. 59 (2018) 287–287. http://jnm.snmjournals.org/content/59/supplement_1/287.abstract.

[71] A. Zwanenburg, M. Vallières, M.A. Abdalah, H.J.W.L. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R.J. Beukinga, R. Boellaard, M. Bogowicz, L. Boldrini, I. Buvat, G.J.R. Cook, C. Davatzikos, A. Depeursinge, M.-C. Desseroit, N. Dinapoli, C.V. Dinh, S. Echegaray, I. El Naqa, A.Y. Fedorov, R. Gatta, R.J. Gillies, V. Goh, M. Götz, M. Guckenberger, S.M. Ha, M. Hatt, F. Isensee, P. Lambin, S. Leger, R.T.H. Leijenaar, J. Lenkowicz, F. Lippert, A. Losnegård, K.H. Maier-Hein, O. Morin, H. Müller, S. Napel, C. Nioche, F. Orlhac, S. Pati, E.A.G. Pfaehler, A. Rahmim, A.U.K. Rao, J. Scherer, M.M. Siddique, N.M. Sijtsema, J. Socarras Fernandez, E. Spezi, R.J. H.M. Steenbakkers, S. Tanadini-Lang, D. Thorwarth, E.G.C. Troost, T. Upadhaya, V. Valentini, L.V. van Dijk, J. van Griethuysen, F.H.P. van Velden, P. Whybra, C. Richter, S. Löck, The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping, Radiology. (2020) 191145. doi: 10.1148/radiol.2020191145.

[72] W.E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods, Biostatistics 8 (2007) 118–127, https://doi.org/10.1093/biostatistics/kxj037.

[73] F. Orlhac, F. Frouin, C. Nioche, N. Ayache, I. Buvat, Validation of A method to compensate multicenter effects affecting CT radiomics, Radiology 291 (2019) 53–59, https://doi.org/10.1148/radiol.2019182023.

[74] F. Orlhac, S. Boughdad, C. Philippe, H. Stalla-Bourdillon, C. Nioche, L. Champion, M. Soussan, F. Frouin, V. Frouin, I. Buvat, A postreconstruction harmonization method for multicenter radiomic studies in PET, J. Nucl. Med. 59 (2018) 1321–1328, https://doi.org/10.2967/jnumed.117.199935.